Topology-based Feature Definition and Analysis

Gunther H. Weber

Computational Research Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720

Peer-Timo Bremer

Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Box 808, L-422, Livermore, CA 94551

Attila Gyulassy and Valerio Pascucci

Scientific Computing and Imaging Institute, University of Utah, 72 S Central Campus Drive, Salt Lake City, UT 84112

Abstract. Defining high-level features, detecting them, tracking them and deriving quantities based on them is an integral aspect of modern data analysis and visualization. In combustion simulations, for example, burning regions, which are characterized by high fuel-consumption, are a possible feature of interest. Detecting these regions makes it possible to derive statistics about their size and track them over time. However, features of interest in scientific simulations are extremely varied, making it challenging to develop cross-domain feature definitions. Topology-based techniques offer an extremely flexible means for general feature definitions and have proven useful in a variety of scientific domains. This paper will provide a brief introduction into topological structures like the contour tree and Morse-Smale complex and show how to apply them to define features in different science domains such as combustion. The overall goal is to provide an overview of these powerful techniques and start a discussion how these techniques can aid in the analysis of astrophysical simulations.

1. Introduction

Extracting quantitative measurements from scientific data is becoming an increasingly important aspect of visual data analysis. For this purpose, defining and extracting features plays a key role in the analysis process. Examples for features of interest (see Figure 1) include burning regions in combustion simulations, bubbles in the mixing layer of a Rayleigh-Taylor instability or filament structures in a porous medium. Once features are detected and extracted, one can derive quantitative measurements such as feature count (number of burning regions or bubbles) or their size distribution and track their evolution over time. In the following, we concentrate on defining features for scalar fields, i.e., a function that assigns a scalar value to each location in the domain.

For scalar fields, two classes of feature definition are particularly ubiquitous: threshold-based features and gradient-based features. Threshold-based features, such as burning regions in a combustion simulation, which can be identified by thresholding the fuel consumption rate, are based on absolute value. They are also closely related to iso-
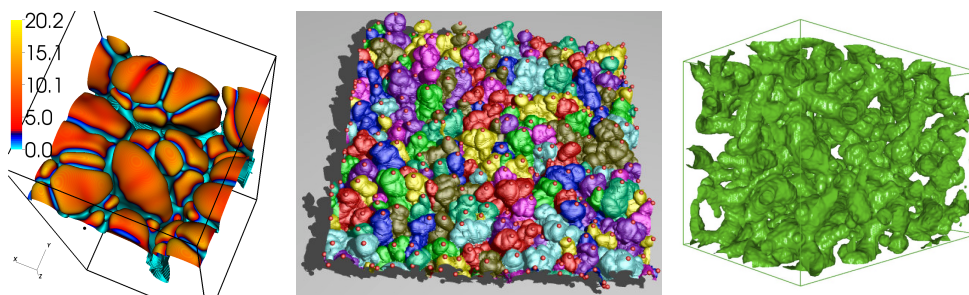
Figure 1.   Features of interest in various scientific data sets. (Left) Burning regions in a combustion simulations. Shown is an isotherm colored by fuel consumption rate. Features of interest are regions of intense burning (orange). (Center) Bubble structure of in the mixing layer of a Rayleigh-Talyor instability simulation. (Right) Filament structure in a porous medium.

surfaces (Lorensen & Cline 1987; Montani et al. 1994; Nielson 2003), which are an extremely versatile and ubiquitous analysis component. An isosurface connects all locations in the domain where the scalar field assumes a specified isovalue. By varying the isovalue and observing changes to the corresponding isosurface, one gains considerable insights into properties of that scalar function. Additionally, isosurfaces arise naturally in data analysis since distinct isosurfaces often have a useful physical interpretation directly related to the application domain. In premixed combustion simulations, an isotherm (isosurfaces of temperature) of the appropriate temperature often is associated with the location of the flame; in molecular simulations, isosurfaces of a certain charge density indicate molecular boundaries; and in fluid simulations, isosurfaces can be used to identify an interface between mixing fluids.

Gradient-based features, on the other hand, consider the vector field obtained by taking the gradient in each location of a scalar field. The image watershed, which considers an image as topographic relief and segments it into watersheds, i.e., regions where a following the gradient reaches the same local minimum, is a gradient-based feature often used in image processing. Examples of gradient-based features in scientific simulations include bubble structures or filament structures in porous media, see Figure 1. Both threshold- and gradient-based features can be related to the topology isosurfaces (Section 2). Using these approaches enables the analysis of a wide range of scientific data sets (Section 3).

2.   Topology of Isosurfaces and Topological Structures

In the context of scalar field analysis, topology-based methods characterize fundamental isosurface properties: the number of connected isosurface components and the genus of an isosurface (i.e., the number of independent tunnels or "holes" in the surface). Morse theory (Milnor 1963) provides insight into the topological evolution isosurfaces of scalar functions defined on manifolds: Isosurface topology changes at distinct isolated *critical points* in the scalar field that can be identified and analyzed analytically as locations where the gradient vanishes. The type of a critical point can be identified by examining the Eigenvalues of the Hessian. Defining critical points in terms of combinatorial criteria (Banchoff 1967, 1970) leads to a numerically more robust way of

identifying them. Critical points are valuable navigation aids when exploring a data set using isosurface extraction (Weber et al. 2002) or when constructing a transfer function for direct volume rendering (Fujishiro et al. 1999, 2000).

Critical points identified by the Morse theory constructions provide hints about the location of interesting behavior in a scalar field. However, as a set of individual disconnected points they are only of limited use to decipher global structures in a data set. This limitation is further aggravated by the fact that critical points can correspond to features at different scales. In a molecular simulation, for example, critical points can correspond to features at the atomic level (such as locations of nuclei) and features at the molecular level (such as bonds between atoms). It is also possible that some critical points arise from numerical "noise" or computational artifacts. Ideally, the analysis should adaptively adjust to focus attention toward details at a given scale, e.g., the molecular scale.

It is extremely useful to consider topological structures that "connect" critical points and relate them to one another. Knowledge of such structures makes it possible to define topological simplification operations that remove features at scales that are not of interest or features due to noise. The *Reeb graph* (Reeb 1946) is one of these structures; it expresses the evolution of individual contours as a graph that is defined by these critical points and their relationships. Consider a differentiable function, f, on a compact manifold (such as a surface embedded in three dimensions). A level set of f consists of all points on the manifold at a specified value. To define the Reeb graph, we consider individual connected components, or contours, of the level sets of f. To obtain the Reeb graph, one assumes that two points on the manifold are equivalent when they belong to the same contour. The quotient space of the manifold with respect to this equivalence relation is the Reeb graph, which essentially is a skeleton of the manifold. Edges of the Reeb graph correspond to families of topologically equivalent contours, and nodes correspond to critical points where the number of contours change.

For dependent scalar variables given on a simply connected simulation domain, the Reeb graph is always a tree structure, called a *contour tree* (Boyell & Ruston 1963). Like the Reeb graph, the contour tree captures the topological evolution of an isosurface as the isovalue varies. Contour tree nodes are the critical points of the data set where the number of contours (connected components) changes. Nodes of degree one (leaves of the tree) correspond to extrema, where contours are created or destroyed. Interior nodes of degree three or higher correspond to "saddles," where two or more contours merge (or a single contour splits into multiple disconnected contours). Arcs of the contour tree represent contours between critical points, i.e., contours that do not change topology (with the exception of genus changes) as the isovalue varies between critical values. Both Reeb graph, and its simpler version, the contour tree, support characterizing threshold-based features.

The *Morse complex* and the *Morse-Smale complex* (Edelsbrunner et al. 2003b,a; Bremer et al. 2004) comprise alternate means of relating critical points to each other, and provide more insight in gradient-based features. Assuming the function on the manifold is differentiable, and thus the gradient is defined at each location, it is possible to start at any location in the domain and follow a gradient line either to its origin or its destination, i.e., follow a line of steepest descent or ascent from that point. This line of steepest ascent/descent will end in a local maximum/minimum. The Morse complex is the segmentation of the domain into ascending or descending manifolds of the critical points. An ascending/descending manifold of a maximum/minimum is the union of

all locations in the domain for which the path of steepest ascent/descent ends at that maximum/minimum. Superimposing ascending and descending manifolds results in the Morse-Smale complex.

Using a Reeb graph, contour tree or Morse-Smale complex, it is possible to rank topological features and simplify the global topological structure (Carr et al. 2004; Gyulassy et al. 2005). This simplification supports further "boiling down" the data to essential features and extracting a simpler, overall coarser structure.

3.   Applications

Using contour tree simplification, it is possible to manipulate isosurfaces in terms of fewer, important connected components (Carr & Snoeyink 2003; Carr et al. 2004), represent the contour tree in a hierarchical layout that makes it possible to explore the global structure of a data set and subsequently obtain a more detailed topology representation in regions of interest (Pascucci et al. 2009), and present alternate representations that represent topological information in a more intuitive fashion (Weber et al. 2007). Using topological structures also has proven useful in analyzing scientific simulation data.

It is possible to apply the Morse-Smale analysis framework to study the influence of turbulence on the evolution of the cellular burning patterns observed in lean premixed hydrogen-air flames. Extracting an isotherm (isosurface of temperature) and segmenting it based on local fuel consumption rate yields a representation of individual burning cells that evolve in time. Since the network of burning cells may depend strongly on the threshold that determines the segmentation, it is crucial to examine the statistical characterization of these flame cells as a function of the threshold value to ensure that the conclusions drawn from the analysis do not depend on the detailed values of this arbitrary parameter (Day et al. 2009). The Morse complex (Bremer et al. 2009a) provides a framework for evaluating the segmentation for a range of thresholds and determining the stability of these arbitrary parameters. Based on a fixed threshold derived from these parameter studies, it is then possible to track burning regions and observe their evolution over time. The Reeb graph of the 3D surface that is swept by the boundaries of burning regions in 4D space time efficiently represents a complete tracking graph (Weber et al. 2009), see Figure 2. A more recent approach uses join trees, which arise in the elegant and efficient contour tree algorithm proposed by Carr et al. (2003), to provide a summary representation of all threshold-based segmentations in a combustion simulation (Bremer et al. 2009b). Using this representation, augmented with derived measurements like burning region volume, it is possible to perform a wide variety of analyses on a greatly reduced amount of data.

The use of topological structures, such as Reeb graph, contour tree and Morse-Smale complex, has also proved useful in a wide variety of other application areas. For example, these structures make it possible to identify and characterize features, such as bubbles in fluid mixing simulations (Laney et al. 2006), which are difficult to characterize using traditional feature detection methods. More recent work used the Morse-Smale complex for calculating clean distance fields for porous solids hit by a projectile (Gyulassy et al. 2007) and using the results to calculate measures such as filament density. Other topological structures, such as Jacobi sets, allow relating critical points between different functions/simulations (Edelsbrunner et al. 2004). Recently,
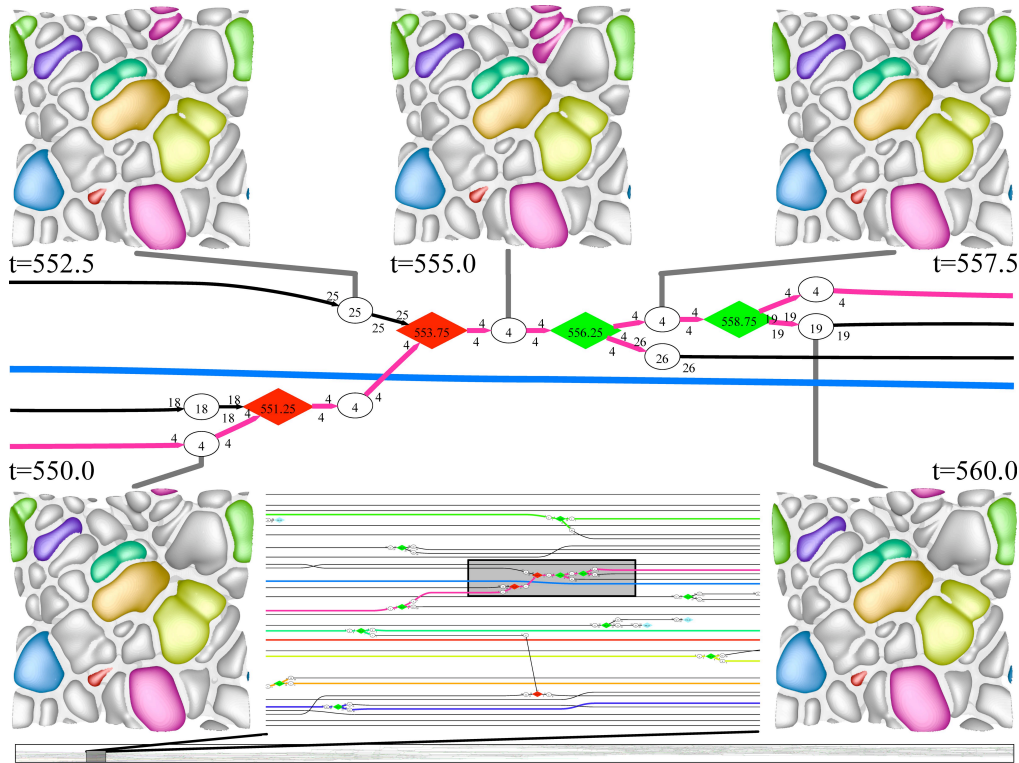
Figure 2. Subsection of the tracking graph of a combustion simulation that is compressed to only show two merge and two split events. Arc color corresponds to region color in corresponding segmentations. Round nodes correspond to cells explicitly segmented by the Morse complex, diamonds to topological events between time steps. Red signifies a merge, green a split, and turquoise a birth/death event. The figure shows three zoom levels of the graph: graph for the entire simulation (bottom graph), a portion corresponding to several subsequent time steps (middle graph) and a zoom into the two merge and split events annotated with colored segmentations of the isotherm.

Bremer et al. (Bremer et al. 2007) gave an overview of topology-based feature definition and feature tracking.

References

Banchoff, T. F. 1967, Differential Geometry, 3, 257
— 1970, American Mathematical Monthly, 77, 457
Boyell, R. L., & Ruston, H. 1963, in Proceedings of the 1963 Fall Joint Computer Conference (IEEE), 445

Bremer, P.-T., Bringa, E. M., Duchaineau, M. A., Gyulassy, A. G., Laney, D., Mascarenhas, A., & Pascucci, V. 2007, Journal of Physics: Conference Series (Proceedings of SciDAC 2007), 78. Doi:10.1088/1742-6596/78/1/012007

Bremer, P.-T., Edelsbrunner, H., Hamann, B., & Pascucci, V. 2004, IEEE Transactions Visualization and Computer Graphics, 10, 385

Bremer, P.-T., Weber, G. H., Pascucci, V., Day, M. S., & Bell, J. B. 2009a, Transactions on Visualization and Computer Graphics. Accepted with minor revisions

Bremer, P.-T., Weber, G. H., Tierny, J., Pascucci, V., Day, M. S., & Bell, J. B. 2009b, in Proceedings of the 5th IEEE International Conference on e-Science (Oxford, UK), 247

Carr, H., & Snoeyink, J. 2003, in Data Visualization 2003 (Proceedings VisSym 2003) (New York, NY: ACM Press), 49

Carr, H., Snoeyink, J., & Axen, U. 2003, Computational Geometry – Theory and Applications, 24, 75

Carr, H., Snoeyink, J., & van de Panne, M. 2004, in IEEE Visualization 2004 (IEEE), 497

Day, M. S., Bell, J. B., Bremer, P.-T., Pascucci, V., & Vince Beckner, M. L. 2009, Combustion and Flame, 156, 1035. Doi:10.1016/j.combustflame.2008.10.029

Edelsbrunner, H., Harer, J., Natarajan, V., , & Pascucci, V. 2004, in Proceedings of IEEE Visualization 2004, 275

Edelsbrunner, H., Harer, J., Natarajan, V., & Pascucci, V. 2003a, in Proceedings of the 19th ACM Symposium on Computational Geometry, 361

Edelsbrunner, H., Harer, J., & Zomorodian, A. 2003b, Discrete & Computational Geometry, 30, 87

Fujishiro, I., Azuma, T., & Takeshima, Y. 1999, in IEEE Visualization '99 (Los Alamitos, California: IEEE Computer Society Press), 467

Fujishiro, I., Takeshima, Y., Azuma, T., & Takahashi, S. 2000, IEEE Computer Graphics and Applications, 20, 46

Gyulassy, A., Natarajan, V., Duchaineau, M., Pascucci, V., Bringa, E. M., Higginbotham, A., & Hamann, B. November/December 2007, IEEE Transactions on Computer Graphics, Proceedings of Visualization 2007, 13

Gyulassy, A., Natarajan, V., Pascucci, V., Bremer, P.-T., & Hamann, B. 2005, in IEEE Visualization 2005, 535

Laney, D., Bremer, P.-T., Macarenhas, A., Miller, P., & Pascucci, V. 2006, IEEE Transcations on Visualization and Computer Graphics, 12, 1053

Lorensen, W. E., & Cline, H. E. 1987, Computer Graphics (Proceedings of ACM SIGGRAPH 87), 21, 163

Milnor, J. W. 1963, Morse Theory (Princeton, New Jersey: Princeton University Press)

Montani, C., Scateni, R., & Scopigno, R. 1994, The Visual Computer, 10, 353

Nielson, G. M. 2003, IEEE Transactions on Visualization and Computer Graphics, 9, 341

Pascucci, V., Cole-McLaughlin, K., & Scorzelli, G. 2009, in Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration (Springer-Verlag), 19

Reeb, G. 1946, Comptes Rendus de l'Acadèmie des Sciences de Paris, 222, 847

Weber, G., Bremer, P.-T., Bell, J., Day, M., & Pascucci, V. 2009, in Topology-based Methods in Visualization (Proc. TopoInVis 2009) (Springer Verlag). In review

Weber, G. H., Bremer, P.-T., & Pascucci, V. November/December 2007, IEEE Transactions on Computer Graphics, Proceedings of Visualization 2007, 13, 1416. LBNL-63763

Weber, G. H., Scheuermann, G., Hagen, H., & Hamann, B. 2002, in IEEE Visualization 2002, IEEE (Los Alamitos, California: IEEE Computer Society Press), 171

Disclaimer